# King's Research Portal

*DOI:*
10.1088/1741-2560/13/4/046011

*Document Version*
Peer reviewed version



*Citation for published version (APA):*
Kamavuako, E. N., Scheme, E. J., & Englehart, K. B. (2016). Determination of optimum threshold values for EMG time domain features; a multi-dataset investigation. Journal Of Neural Engineering, 13(4), 046011. DOI: 10.1088/1741-2560/13/4/046011

# Determination of Optimum Threshold Values for EMG Time Domain Features; a Multi-Dataset Investigation

Ernest Nlandu Kamavuako[1*], Erik Justin Scheme[2], Kevin Brian Englehart[2]

[1]Center for Sensory-Motor Interaction, Department of Health Science and Technology, Aalborg University, Fredrik bajers vej 7 D3, 9220 Aalborg, Denmark.

[2]Institute of Biomedical Engineering, Department of Electrical and Computer Engineering, University of New Brunswick, 15 Dineen Dr., D36 Head Hall, Fredericton, E3B 5A3 NB, Canada

**Keywords**: Electromyogram, pattern recognition, Time domain features, slope sign changes, zero crossing.

*Corresponding Author: Ernest N. Kamavuako, Center for Sensory-Motor Interaction (SMI), Department of Health Science and Technology, Aalborg University, Fredrik Bajers Vej 7 D3, DK-9220 Aalborg, Denmark, Tel: +45 99402461. E-mail: enk@hst.aau.dk

## ABSTRACT

*Objective:* For over two decades, Hudgins' set of time domain features have extensively been applied for classification of hand motions. The calculation of Slope Sign Change (SSC) and Zero Crossing (ZC) features uses a threshold to attenuate the effect of background noise. However, there is no consensus on the optimum threshold value. In this study, we investigate for the first time the effect of threshold selection on the feature space and classification accuracy using multiple datasets.

*Approach:* In the first part, four datasets were used, and classification error (CE), separability Index, scatter matrix separability criterion, and cardinality of the features were used as performance measures. In the second part, data from eight classes were collected during two separate days with two days in between from eight able-bodied subjects. The threshold for each feature was computed as a factor (R = 0:0.01:4) times the average root mean square of data during rest. For each day, we quantified classification error for R = 0 ($CE_{r0}$) and minimum error ($CE_{best}$). Moreover, a cross day threshold validation was applied where, for example, CE of day two ($CE_{odt}$) is computed based on optimum threshold from day one and vice versa. Finally, we quantified the effect of the threshold when using training data from one day and test data of the other.

*Main results*: All performance metrics generally degraded with increasing threshold values. On average, $CE_{best}$ (5.26 ± 2.42 %) was significantly better than $CE_{r0}$ (7.51 ± 2.41 %, P = 0.018), and $CE_{odt}$ (7.50 ± 2.50 %, P = 0.021). During the two-fold validation between days, $CE_{best}$ performed similar to $CE_{r0}$. Interestingly, when using the threshold values optimized per subject from day one and day two respectively, on the cross-days classification, the performance decreased.

*Significance*: We have demonstrated that threshold value has a strong impact on the feature space and that an optimum threshold can be quantified. However, this optimum threshold is highly data and subject driven and thus do not generalize well. There is a strong evidence that R = 0 (no trheshold) provides a good trade-off between system performance and generalization. These findings are important for practical use of pattern recognition based myoelectric control.

# I. INTRODUCTION

For individuals with upper-limb amputation, myoelectric signals used for pattern recognition play a key role in advanced control of powered multifunctional prostheses. The success of such a system depends highly on the classification accuracy, which is dependent on the choice of features and classifiers (Huang et al. 2005, Li 2011, Shin et al. 2014). In fact, it has been shown that classification accuracy is more affected by the choice of feature set than by the choice of classifier (Hargrove et al. 2007, Phinyomark et al. 2008). Three-types of features, namely time domain (TD), frequency domain, and time-frequency domain features, are dominant in the literature. TD features have been widely used in myoelectric classification due to their computational simplicity and because they are easy to implement. TD features are based on signal amplitudes extracted directly from raw EMG data without requiring any transformation. Hudgins' TD feature set was introduced in 1993 and is by far the most commonly used feature set in research and commercial implementation (www.CoAptEngineering.com). These TD features include mean absolute value (MAV), waveform length (WL), slope sign changes (SSC), and zero-crossing (ZC) (Hudgins et al. 1993). SSC and ZC are of interest in the present study, because their computation includes a threshold value as shown in Eq. (1) and (2). Hudgins et al. proposed that a threshold must be included in the ZC and SSC calculation to reduce aberrant zero crossings or slope sign changes resulting from additive noise. According to Hudgins et al., the threshold value is system dependent and in their case, the value was ±10 mV (using a system gain of 5000) Since then, most studies using this set of features employ a threshold; however, after more than 20 years, there is still no consensus on the optimum value.  Furthermore, the effect of this threshold value on the efficacy of EMG features is not well understood.

A limited number of studies have investigated the effect of different threshold values. Interestingly the case of the threshold being equal to zero was not included. Phinyomark et al. (2009), investigated the most suitable threshold value for classifying four motions using surface electromyography (EMG) recorded from the extensor and flexor carpi radialis. Threshold values from 10 to 50 mV (system gain of 1000) were tested for varying signal to noise ratio (SNR). They reported optimal thresholds of 30 mV for SSC and 10 mV for ZC. Furthermore, they reported that the threshold value was gain and instrument dependent. Kamavuako et al. (2013) investigated the threshold value of these features for force estimation; and found that the threshold was data dependent.

Other studies have applied predefined threshold values without investigating the effect of different threshold values. For example, (Zardoshti et al. 1995) used no threshold for ZC; (Phinyomark et al. 2010) used a threshold value of 10 mV for SSC and ZC. In (Phinyomark et al. 2013), the threshold value for SSC and ZC was set to 16 mV. The list is not exhaustive as there are many other studies using Hudgins' TD feature set where threshold is not mentioned. Although it is believed that most studies apply a threshold value for SSC and ZC, there is no clear evidence on why eliminating low-level details in SSC and ZC is helpful; and how the threshold affects the feature space. In fact, computation of MAV and WL does include low-level details. The fact that the threshold value presumably depends on the instrument and dataset is alarming for the clinical use of pattern recognition based systems. Furthermore, in previously mentioned studies, robustness over time and generalization for multiple datasets has not been characterized, since investigations were based on acute recordings from single day (session) experiments. Robustness and the ability to generalize is essential for practical use of pattern recognition-based control systems.

In the present study, we use multiple datasets to revisit Hudgins' TD feature set by investigating how thresholding SSC and ZC affects the feature space and classification performance. The aim was to test whether a global optimum threshold factor (across datasets) can be found. Furthermore, we investigated the robustness of optimized thresholds over two days.

## II. METHODS

### *A. Features*

For this investigation, the focus is on Hudgins' features that capture frequency information such as *ZC* and *SSC*. Hudgins et al. (1993) proposed that a threshold (ε) condition be applied for both features to avoid including contributions from background instrumentation and quantization noise. *ZC* measures the number of times the signal crosses zero relates to the frequency content of the signal**Error! Reference source not found.**. It is defined as given in Eq. (1). SSC measures the number of times the sign changes in the slope of the signal, also capturing frequency information. The definition of SCC is provided by Eq. (2).

$$ZC = \sum_{k=1}^{N-1} [(x_k \cdot x_{k+1} < 0) \cap (|x_k - x_{k+1}| > \epsilon)] \quad (1)$$

$$SSC = \sum_{k=2}^{N-1} f\,[(x_k - x_{k-1}) \cdot (x_k - x_{k+1})] \quad (2)$$

$$f(x) = \begin{cases} 1, & if\ x > \epsilon \\ 0, & otherwise \end{cases}$$

It should be noted that we have proposed a modification of the threshold condition of SSC from $x \geq \epsilon$ to $x > \epsilon$ in order to avoid counting DC values. This is a small modification, but necessary for the SSC to maintain valuable signal information at $\epsilon = 0$.

In order to make the threshold generic for any system gain, it is defined as a factor ($R$) times the average (across channels) root mean square value of the EMG signal at rest (during no contraction) (Eq. 3). In this study $R$ ranges from 0 to 4 with a step of 0.01.

$$\epsilon = R * \sqrt{\frac{1}{N}\sum_{j=1}^{N}(x_{NM}[j])^2} \tag{3}$$

where $x_{NM}[j]$ are the samples of the signal at rest, and $N$ is the total number of samples.

The methodology of this study is divided into two stages. In the first stage, four datasets are used to investigate the threshold effect on both the feature space and on the overall performance. In the second stage, data were collected on two separate days (within a week) to evaluate the robustness of selective (optimized) threshold values and the inter-day generalization.

### *B. Multi dataset investigation*

#### *1) Experimental procedures*

For this investigation, we used data recorded using two different experimental protocols. In the first protocol, the experiments were conducted on nine able-bodied subjects (6 male/3 female, age range: 19 - 26 yrs). The procedures were in accordance with the Declaration of Helsinki and approved by the University of New Brunswick's research ethics board. Subjects provided their written informed consent prior to the experimental procedures. The subjects had no history of upper extremity or other musculoskeletal disorders. Surface EMG was recorded from the following muscles: flexor carpi radialis, flexor digitorum superficialis, extensor carpi radialis and extensor digitorum communis using four bipolar electrodes (Duo-trode Ag-Ag/Cl, Myotronics, Inc. USA). Surface EMG signals were analog bandpass filtered between 10 – 500 Hz. All signals were amplified (AnEMG12, OTbioelletronica, Torino, Italy), A/D converted using 16 bits (NI-DAQ USB-6259), and sampled at 10 kHz. Subjects were prompted to elicit comfortable and sustainable contractions corresponding to seven classes of motion; wrist flexion (WF), wrist extension (WE), hand close (HC), hand open (HO), key grip (KG), pinch grip (PG) and no motion (NM). Four repetitions of three seconds were collected for each motion, during which the unconstrained subjects held a medium level contraction to capture signals a steady state and during ramp contractions (from rest to medium contraction). These data were divided to form two datasets; one for steady state and the other for ramp contractions.

In the second protocol, previously collected data, as reported in Kamavuako et al. 2014, were used. Briefly, the experiments were conducted on nine able-bodied subjects (7 male/2 female, mean age: 26.5 years old). Intramuscular EMG was recorded using two bipolar wire electrodes from the supinator teres and pronator teres muscles. Intramuscular signals were analog bandpass filtered between 0.1 and 4.4 KHz. Surface EMG was recorded using four bipolar electrodes (Ag-Ag/Cl, Red Dot™, Canada) placed over the flexors and extensors compartments respectively. On the flexor side, one channel was placed over the Flexor Digitorum Profondus and Flexor Carpi Ulnaris. On the extensor side, the other channel was placed over the Extensor

Carpi Ulnaris and the Extensor Digitorum Communis. These two surface channels were combined with the two intramuscular channels to create a combined EMG dataset. The combined data EMG dataset were used here because the same features were applied [Kamavuako et al. 2014]; and this combination showed superior performance compared to surface EMG. Surface EMG signals were analog bandpass filtered between 10 – 500 Hz. All signals were amplified (AnEMG12, OTbioelletronica, Torino, Italy), A/D converted using 16 bits (NI-DAQ USB-6259), and sampled at 10 kHz. A reference electrode was placed close to the carpus. Subjects were prompted to elicit comfortable and sustainable contractions corresponding to nine classes of motion; wrist supination (*WS)*, wrist pronation (*WP)*, *HO*, *HC*, simultaneous *WS*+*HO*, *WS*+*HC*, *WP*+*HO*, *WP*+*HC* and NM. Four repetitions of two seconds were collected for each motion, during which the unconstrained subjects held a medium level contraction to capture signals at a steady state. The collected data provided two datasets using surface and combined EMG. Thus from the two protocols described above four datasets were extracted and used for this investigation.

*2) Data analysis*

EMG signals were digitally high-pass filtered (third order Butterworth filter) with a cutoff frequency at 20 Hz to attenuate movement artifacts. Four time-domain features were extracted from incrementing (by 25 ms) analysis windows of 200 ms in duration. Data analysis was carried out first by investigating ZC and SSC alone as features and then by combining SSC and ZC each with the MAV and WL features. Thus, only three features were used at a time. For both cases, the following measures were used on each dataset to evaluate the effect of varying values of *R*.

1. Classification Error (CE): For each *R-value,* CE was quantified using four-fold validation procedure with Linear Discrimimant Analysis as the classifier. To quantify the effect of the threshold on SSC and ZC, CE was computed using the optimum *R*-value per subject ($CE_{best}$), error at best *R-value* for the population ($CE_{avg}$), error at *R* = zero ($CE_{r0}$) and for maximum error ($CE_{worst}$).
2. Separability Index (SI): This metric was introduced by Bunderson and Kuiken (2012) as a measure for interclass distances. It is defined as one-half the Mahalanobis distance from the centroid of the training ellipsoid of class *j* to the centroid of the training ellipsoid of the nearest class, averaged over all classes. A larger SI indicates classes that are more distinct.
3. Scatter matrix separability criterion (SMSC): SMSC is a frequently used measure of class separability (Wang 2008, Wang and Chan 2002). The scatter matrices include the within-class scatter matrix, $S_w$, and the between-class scatter matrix, $S_b$. A smaller $S_w$ and larger $S_b$ means larger class separability. It is computed as the trace of the ratio between $S_b$ and $S_w$.
4. Cardinality of SSC and ZC: Cardinality of a set is a measure of the number of distinct values. This is used in this study as a measure of dynamic range. We chose not to use the usual definition of dynamic

range (ratio between maximum and minimum value) because SSC and ZC are counts based features. Cardinality of EMG samples has been recently proposed as a robust feature itself for myoelectric control (Ortiz-Catalan 2015), however here we use it to quantify how the dynamic range of the SSC and ZC features changes with the threshold.

5. Effect of noise for classification: here we investigated the effect of background noise on classification when a threshold is applied and when no threshold is applied. For this investigation, surface EMG dataset (second protocol) was artificially contaminated with noise at different signal-to-noise ratios (SNR). The procedure was as followed: white Gaussian noise (WGN) was generated at SNR between -60 to 60 dB with a step of 5 dB. The generated noise was first bandlimited between 10 and 500 Hz and then added to the EMG signals; however the reported SNR values are still based on WGN. Ten realizations of WGN were generated on channel basis to compensate for the randomness on classification.

### 3) *Statistics*

Statistical analysis was carried on classification error alone. The results of all datasets were aggregated prior to the analysis to allow the use of one-way repeated measures analysis of variance (ANOVA) in order to quantify the effect of threshold on SSC and ZC both when used alone and together with MAV and WL. The factors were $CE_{r0}$, $CE_{best}$, $CE_{avg}$, and $CE_{worst}$. P-values less than 0.05 were considered significant. The Bonferroni–Dunn adjustment was used for multiple comparisons. Results are given as mean ± standard error.

## C. *Cross day robustness of threshold values*

### 1) *Experimental procedures*

Eight healthy subjects participated in the experiment (five males and three females, 25 ± 1 years old). The experiment was approved by the University of New Brunswick’s research ethics board, and carried out in accordance with the Declaration of Helsinki. The subjects gave informed consent prior to the experiment. Surface EMG was recorded from five channels (EMG-USB2, OT Bioelettronica) using Ag/AgCl electrodes (Ambu Neuroline 720) placed on the pronator teres, flexor digitorum superficialis, flexor carpi radialis, extensor digitorum, and extensor carpi radialis muscles. All signals were analog filtered between 10-500 Hz, amplified (AnEMG12, OTbioelletronica, Torino, Italy) 2000 times, A/D converted using 16 bits (NI-DAQ USB-6259), and sampled at 2 kHz. Seven hand movements were selected based on frequent use in activities of daily living: HO, HC, WF, WE, WS, WP, PG. Furthermore, recordings of NM were performed. Surface EMG data were collected from forearm muscles during two separate days with two days in between collections. Each day, four repetitions were collected for each movement. Each movement was recorded for three seconds and Surface EMG signals were only collected after the subject had reached a steady-state contraction of the required movement. The subjects were instructed to make medium, constant force contractions to the best of their ability. The electrode positions were marked to maintain the same electrode locations on different days.

*2) Data analysis*

EMG signals were digitally high-pass filtered (3rd order Butterworth filter) with a cutoff frequency at 20 Hz to attenuate any movement artifacts. Four time-domain features were extracted from incrementing (by 25 ms) analysis windows of 200 ms in duration. For each day, CE was quantified for each threshold value to investigate whether the relationship between CE and *R* is the same for the two days. Furthermore, the threshold associated with the lowest error for each subject was recorded. In brief, for each day, classification error for *R* = 0 ($CE_{r0}$), and minimum error ($CE_{best}$) with associated best threshold ($R_{best}$) and worst error ($CE_{worst}$) were computed. Moreover, a cross day threshold validation was applied where the CE of each day was computed using the optimum threshold determined from the other day ($CE_{odt}$). Finally, the effect of threshold when using training data from one day and test data of the other day was quantified using a two-fold validation across both days. For these purposes, all combinations of *R* (0:001:4) for SSC and ZC were investigated.

*3) Statistics*

Two-way repeated measures analysis of variance (ANOVA) with factors CE and days was used to compare $CE_{best}$, $CE_{r0}$, $CE_{worst}$ and $CE_{odt}$. The Bonferroni–Dunn adjustment was used for multiple comparisons. P-values less than 0.05 were considered significant.

## III. RESULTS

### *A. Multi dataset investigation*

Figures 1 and 2 depict the effect of increasing threshold on the performance and feature space for SSC and ZC respectively when used alone. The general view of Figure 1a and 2a, indicates that the classification error increases with increasing threshold for SSC, while ZC has a minimum that is not located at R = 0. This is confirmed by both measures of class separability SI and SMSC. Although the maximum SI depends on the dataset, the ensemble average shows a clear advantage of using very low threshold values. SMSC seems to follow the error better than SI. Figure 1d and 2d reconfirm the findings as the dynamic range of SSC and ZC (quantified through cardinality) is reduced with increasing threshold values, especially SSC.

[Figure 1 and 2 about here]

Figure 3 summarizes $CE_{r0}$, $CE_{best}$, $CE_{worst}$ and $CE_{avg}$ when using SSC and ZC as single features for classification. It is clear that $CE_{best}$ is significantly lower ($P < 0.001$) than all other measures for both SSC and ZC. However $CE_{r0}$ (25.4 ± 1.75 %) is exactly the same as $CE_{avg}$ for SSC. For ZC, $CE_{r0}$ (33.5 ± 1.80%) was not significantly different from $CE_{avg}$ (27.2 ± 2.47 %) indicating that $R = 0$ may perform similarly to the case when the threshold is optimized for the average performance of the population.

The range of best threshold values, computed based on minimum CE, maximum SI and maximum SMSC is shown in Figure 4. In general, SSC showed a limited range compared to ZC. In both cases, SMSC showed more consistency with CE.

[Figure 3 and 4 about here]

In Figure 5, we present the behaviour of SSC and ZC when each is combined with the two other TD features (MAV and WL) with $\underline{R}$ (0:0.01:4). An improvement was seen for all measures. For SSC, $CE_{best}$ (9.83 ±1.31 %) was not significantly better ($P = 0.18$) than $CE_{r0}$ (10.35 ± 1.32 %). However both were significantly better ($P < 0.01$) than $CE_{worst}$ (14.49 ± 1.59 %). For ZC, $CE_{best}$ (9.97 ±1.33 %) was significantly better ($P < 0.01$) than $CE_{r0}$ (11.65 ± 1.44 %). Furthermore, $CE_{r0}$ performed similarly to $CE_{avg}$ (11.19 ± 1.38 %), with all three being significantly better ($P < 0.001$) than $CE_{worst}$ (14.7 ± 1.60 %).

[Figure 5 about here]

It can be noted that optimizing the threshold can be effective when done on a subject basis for single features. However, SSC and ZC are almost always used in combination with MAV and WL, thus the value added with the optimization in comparison to setting the threshold to zero seems negligible for SSC; especially when optimization is done based on the averaged population. In the next section, we see how the threshold behaves across days if one chooses to optimize it.

In Figure 6, we present the behaviour of classification error at different SNR values. It is seen that, at higher SNR values, the difference between applying a threshold and no threshold is around two percentage points. When noise become prominent, both cases converged to the same performance, supporting the idea of leaving the threshold out of the equation.

[Figure 6 about here]

### *B. Cross day robustness of threshold values*

During the four-fold validation within each day, there was a significant difference ($P < 0.01$) between CE, but no difference ($P = 0.186$) between days was found with no interaction ($P = 0.442$). $CE_{best}$ (5.26 ± 2.42 %) was significantly better than $CE_{r0}$ (7.51 ± 2.41 %, $P = 0.018$), $CE_{worst}$ (11.53 ± 2.62 %, $P < 0.001$) and $CE_{odt}$ (7.50 ± 2.50 %, $P = 0.021$). $CE_{r0}$ was significantly better than $CE_{worst}$ ($P = 0.001$), but performed similar to $CE_{odt}$ ($P = 1.00$) indicating that optimising the threshold on day one and using it for day two provides similar performance as using $R = 0$. The results are summarized in Figure 7.

[Figure 7 about here]

The average RMS of the no motion class data (across channels and subjects) was 0.01 ± 0.001 V and 0.012 ± 0.001 V for day 1 and 2 respectively; not significantly different ($p = 0.204$). However a significant difference between channels was found ($P = 0.02$). Furthermore, no subject had the same optimal *R-value* for both days. We should note that using channel-based thresholding did not improve the results for this dataset.

During the two-fold validation between days, the minimum of the average CE space was 14.90 ± 3.8 % located at $R = 0.06$ and 2.18 for SSC and ZC respectively. This minimum was not different ($P = 15.73$) from the $CE_{r0}$ (16.17 ± 4.05%) obtained at $R = 0$ for both features. Interestingly, when using the threshold values optimized per subject from day one and day two respectively, on the between days classification, the CE increased to 19.90 ± 4.6% and 22.86 ± 5.67% for day one and day two respectively. Similarly fixing a value of $R$ corresponding to a threshold of 10 mV for all subjects as suggested by Hudgins et al. (1993) performed worse than $R = 0$ (no threshold). This suggests that thresholds optimized on single data sets do not generalize well. The error space of the two days exhibited different shapes.

## IV. Discussion

In the present study, a multi dataset investigation was conducted to investigate how thresholding SSC and ZC affects the feature space and classification performance. The aim was to test whether a global optimum threshold factor (across datasets) can be found. Furthermore, we investigated the robustness of optimized thresholds over two days.

For the multiple dataset investigation using SSC and ZC alone, the ensemble average showed a clear advantage of using very low threshold values. For SSC, a minimum classification error was located at $R=0$ that increased with increasing threshold, while the minimum for ZC was not located at the origin. Furthermore results revealed that it is optimal to identify the threshold on a per subject basis. Each subject had a unique global minimum, supporting the initial suggestion made by Hudgins et al. (1993) for reducing aberrant zero crossings or slope sign changes resulting from additive noise. When adding MAV and WL features to the classification, the advantage of optimizating ZC and SSC in comparison to setting the threshold to zero is still evident. However if optimization is done on the average population, then its effect is statitically negligeable. Interestingly, this is the approach that is usually adopted in the literature. We should note that although all four datasets were collected using the same amplifier, the optimum threshold still depends on the dataset and subject.

The usability of pattern recognition systems cannot be assessed using offline single session data. Among different modalities, realtime performance (Lock et al., 2005 and Hargrove et al., 2007b) and system robustness over time (He et al 2015) have been recommended. In the present study we focus on robustness using data recorded over two days. Robustness of SSC and ZC is essential for practical use of pattern recognition-based control systems. In the ideal world, it is desired that classification performance should not be affected by the threshold value. In offline mode and during single session recordings, one can identify the optimum threshold and thereby optimize the classification performance for a specific investigation. However, it is alarming that thresholding is strongly subject specific. Through threshold investigation, the best threshold can be identified offline and applied during control of prosthesis. We also show that application of WGN did affect performance equally for the case of threshold on no threshold respectively.

All investigations indicate that $CE_{best}$ consistently outperforms all other metrics suggesting that the performance is heavily data driven, and it varies by subject and day. The remaining question is whether it is feasible that optimum threshold value corresponding to $CE_{best}$ could be determined using a validation set. This question cannot be answered fully using the available data for this study, however this is the subject of an ongoing investigation that we hope to share in the near future. With two days between recordings, we did not assess when the performance starts to deviate from $CE_{best}$. However, we have shown that optimizing $R$ on day one and using it on day two provided similar performance as $R = 0$. If the threshold was optimized daily, one would need retraining a classifier, which would be difficult to do several times daily or in real time. Furthermore, the value gained may not justify the effort, as the absolute difference between $CE_{best}$ and $CE_{r0}$ seems small in this study and may not have practical significance.

In the ideal world of pattern recognition, it is desired to have a system in which daily training is not necessary. This would mean that training data recorded in day one can be used on subsequent days. Although our investigation is limited to two days, we have shown that even subject dependent optimum thresholds do not generalize well. $R = 0$ does not provide the optimum performance, but it is a good tradeoff between performance and generalization.

This study investigated for the first time, the effect of thresholding SSC and ZC on the feature space and the classification performance. However, it should be noted that all investigations are carried out offline; thus, whether the difference between $CE_{best}$ and $CE_{r0}$ is practically relevant can only be assessed with patients controlling a device. Nevertheless, a theoretical investigation such as this present work is usually meant to pave the way for more practical and clinically oriented experiments.

## V. CONCLUSION

The aim of this experiment was to investigate the effect on thresholding TD features on the feature space and classification accuracy. We have demonstrated that threshold value has a large impact on the feature space and that an optimum threshold can be quantified. However, this optimum threshold is highly data and subject driven and thus do not generalize well. Results showed a strong evidence that keeping the threshold equals zero provides a good trade-off between system performance and generalization.

**Acknowledgement**

The authors declares that there is no conflict of interest. The authors wish to thank Julie Gade and Rosa Hugosdottir for their contribution during data collection for the two days experiment.

## REFERENCES

Bunderson and Kuiken (2012). Nathan E. Bunderson and Todd A. Kuiken, Quantification of Feature Space Changes With Experience During Electromyogram Pattern Recognition Control. IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING, VOL. 20, NO. 3, 2012.

Hargrove, L. J., Englehart, K., & Hudgins, B. "A comparison of surface and intramuscular myoelectric signal classification". Biomedical Engineering, IEEE Transactions on, vol. 54, no. 5, pp. 847-853, 2007.

Hargrove L, Losier Y, Lock B, Englehart K, Hudgins B. A Real-Time Pattern Recognition Based Myoelectric Control Usability Study Imlemented in a Virtual Environment. Proceedings of EMBS Conference , Lyon, Aug. 2007b.

He J, Zhang D, Jiang N, Sheng X, Farina D, Zhu X. User adaptation in long-term, open-loop myoelectric training: implications for EMG pattern recognition in prosthesis control. J Neural Eng. 2015, 12(4):046005. doi: 10.1088/1741-2560/12/4/046005

Huang, Y., Englehart, K.B., Hudgins, B., & Chan, A. D."A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses". Biomedical Engineering, IEEE Transaction on, vol. 52, no. 11, pp. 1801-1811, 2005.

Hudgins, B., Parker, P., & Scott, R. N. "A new strategy for multifunction myoelectric control". Biomedical Engineering, IEEE Transactions on, vol. 40, no. 1, pp. 82-94, 1993.

Kamavuako, E.N., Scheme, E.J., Englehart, K. B. & Hudgins, B. S. "Global features for the estimation of discharge rate from intramuscular EMG". Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on, pp. 182-185, 2013.

Kamavuako EN; Scheme, Erik J.; Englehart, Kevin B. Combined surface and intramuscular EMG for improved real-time myoelectric control performance. In: Biomedical Signal Processing and Control, Vol. 10, No. 1, 2014, p. 102–107.

Li, G. (2011). Electromyography pattern-recognition-based control of powered multifunctional upper-limb prostheses. INTECH Open Access Publisher, 2011.

Lock B, Englehart K, Hudgins B. Real-Time Myoelectric Control in a Virtual Environment to Relate Usability vs. Accuracy. Myoelectric Controls Symposium, New Brunswick, Fredericton, 2005.

Phinyomark, A., Limsakul, C. & Phukpattaranont, P. "EMG feature extraction for tolerance of white Gaussian noise". Proceedings of the International Workshop and Symposium on Science and Technology, pp. 178-183, 2008.

Phinyomark, A., Limsakul. C. & Phukpattaranont, P. "A novel feature extraction for robust EMG pattern recognition". JOURNAL OF COMPUTING, VOLUME 1, ISSUE 1, 2009.

Phinyomark, A., Hirunviriya., S., Limsakul, C. & Phukpattaranont, P. "Evaluation of EMG feature extraction for hand movement recognition based on Euclidean distance and standard deviation". Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON), 2010 International Conference on, pp. 856- 860, 2010.

Phinyomark, A., Quaine, F., Charbonnier, S., Serviere, C., Tarpin-Bernard, F., & Laurillau, Y. "EMG feature evaluation for improving myoelectric pattern recognition robustness". Expert Systems with Applications, vol. 40, no. 12, pp. 4832-4840, 2013.

Ortiz-Catalan 2015 Cardinality as a highly descriptive feature in myoelectric pattern recognition for decoding motor volition. Front. Neurosci., 2015 http://dx.doi.org/10.3389/fnins.2015.00416

Shin, S., Tafreshi, R., & Langari, R. "A performance comparison of hand motion EMG classification". Biomedical Engineering (MECBME), 2014 Middle East Conference on. IEEE, pp. 353-356, 2014.

Wang, L. "Feature selection with kernel class separability". Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 30, no. 9, pp. 1534-1546, 2008.

Wang, L. & Chan, K. L. "Learning kernel parameters by using class separability measure". 2002

He et al 2015: He J, Zhang D, Jiang N, Sheng X, Farina D, Zhu X. User adaptation in long-term, open-loop myoelectric training: implications for EMG pattern recognition in prosthesis control. J Neural Eng. 2015 Aug;12(4):

Zardoshti-Kermani, M., Wheeler, B. B., Badie, K. & Hashemi, R. M. "EMG feature evaluation for movement control of upper extremity prostheses". Rehabilitation Engineering, IEEE Transactions on, IEEE, vol. 3, no. 4, pp. 324-333, 1995.

**Captions**

Figure 1: Result obtained for slope sign change (SSC) in different dataset. Plain dark line is the average over all dataset. (a) Classification error computed at different threshold levels for SSC. For some dataset, high threshold values resulted in singular covariance matrices for some classes. Thus, we chose not to report this error for these cases (b) Class separabilty quantified through separabilility index. (c) Separability of the feature space quantified through Scatter matrix separability criterion (SMSC) and (d) the cardinality of SSC alone across classes and channels. Data are normalized for easy visualization. Normalization is performed on subject basis by dividing by the maximum of each measure prior to averaging.

Figure 2: Result obtained for zero crossing (ZC) in different dataset. Plain dark line is the average over all dataset. (a) Classification error computed at different threshold levels for ZC. For some dataset, high threshold values resulted in singular covariance matrices for some classes. Thus, we chose not to report this error for these cases (b) Class separabilty quantified through separabilility index. (c) Separability of the feature space quantified through Scatter matrix separability criterion (SMSC) and (d) the cardinality of ZC alone across classes and channels. Data are normalized for easy visualization. Normalization is performed on subject basis by dividing by the maximum of each measure prior to averaging.

Figure 3: Classification error (CE) for (A) SSC and (B) ZC used as single classification features at different $R$ values. $CE_{r0}$ is the CE when $R$ = 0, CEbest is subject specific minimum error, CEworst is the maximum error and CEavg is the CE obtained when R is optimized based on the population. Stars (*) indicates the statistical significant differences with respect to CEbest.

Figure 4: The range of best R values if optimized based on classification error (CE), Scatter matrix separability criterion (SMSC) and separability index (SI).

Figure 5: Classification error (CE) for (A) SSC and (B) ZC used in combination with Mean absolute value and wavelength features at different $R$ values. $CE_{r0}$ is the CE when $R$ = 0, CEbest is subject specific minimum error, CEworst is the maximum error and CEavg is the CE obtained when $R$ is optimized based on the population. Stars (*) indicates the statistical significant differences with respect to CEbest.

Figure 6: Classification error (CE) at different signal-to-noise ratio (SNR) with application of a threshold (gray line with square) value and no threshold (black line with circles) respectively.

Figure 7: Classification error (CE) for day 1 and day 2 showing minimum CE (CEbest), CE when $R$ = 0 ($CE_{r0}$), the maximum CE (CEworst) and cross day threshold (CEdx). Stars (*) insidcates the degree of significant difference.

**Figure 1**

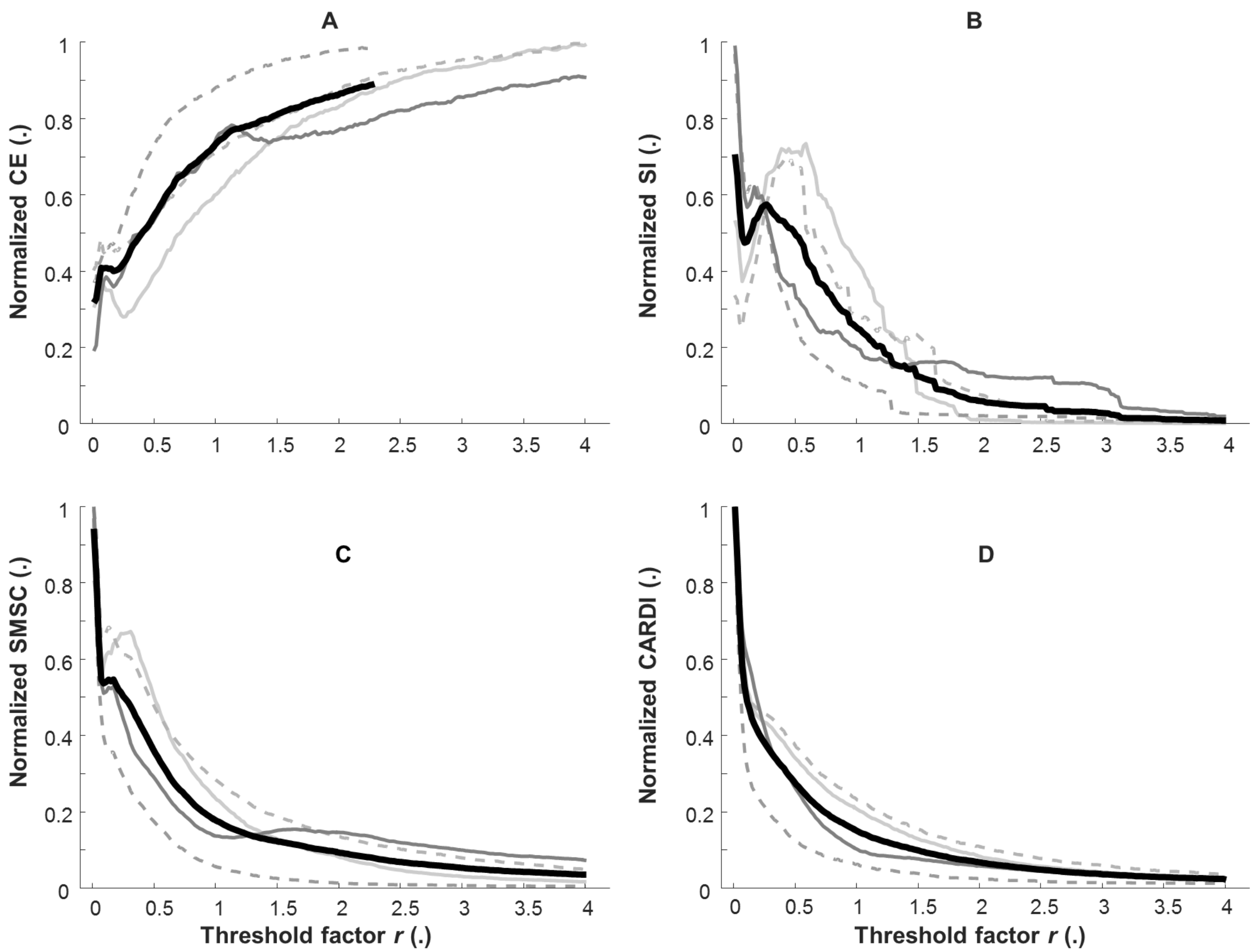

A
B
C
D
Normalized CE (.)
Normalized SI (.)
Normalized SMSC (.)
Normalized CARDI (.)
Threshold factor r (.)
Threshold factor r (.)
0
0.2
0.4
0.6
0.8
1
0.5
1.5
2
2.5
3
3.5
4

**Figure 2**

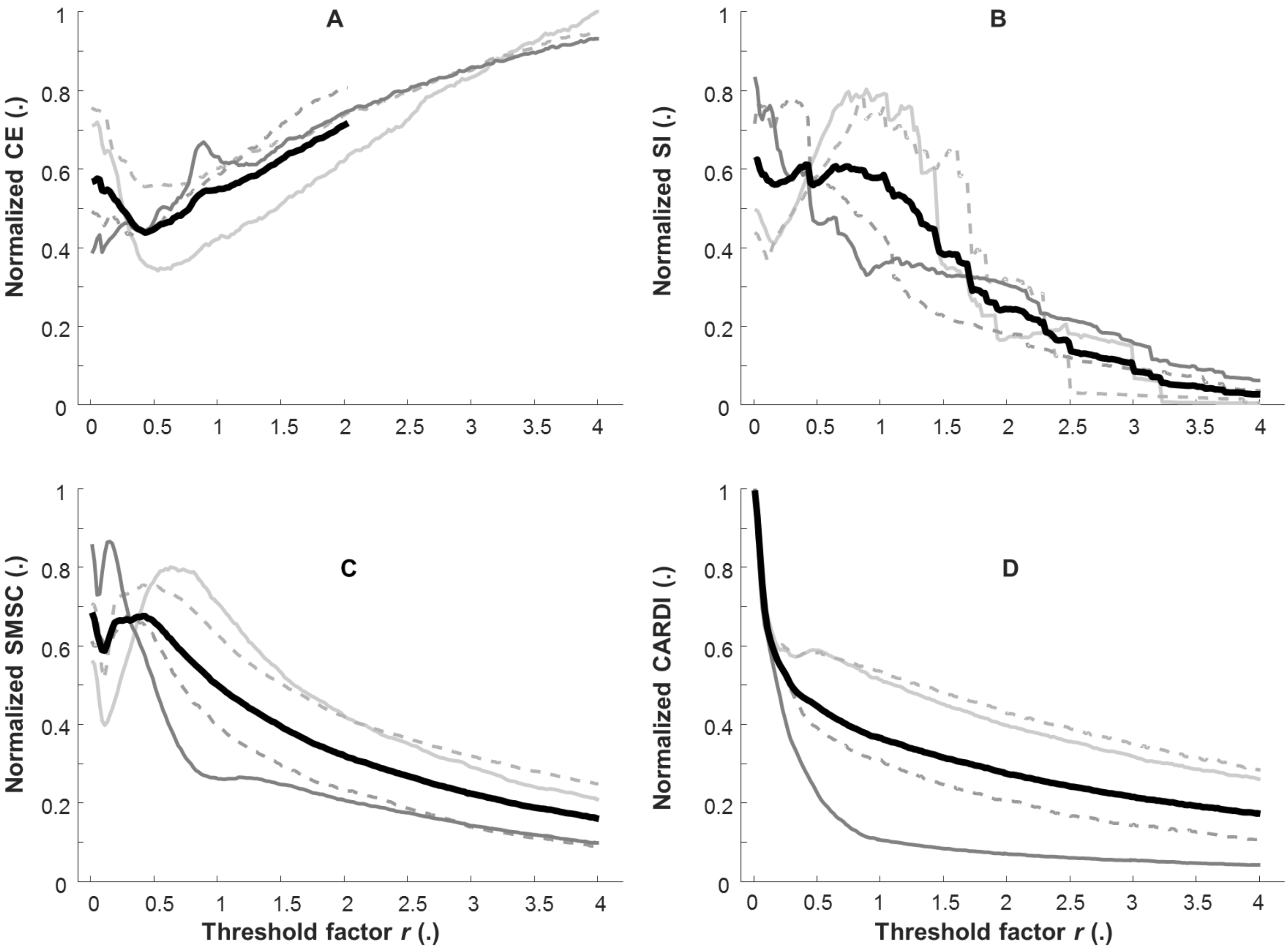

A
B
C
D
Normalized CE (.)
Normalized SI (.)
Normalized SMSC (.)
Normalized CARDI (.)
Threshold factor r (.)
Threshold factor r (.)
0
0.2
0.4
0.6
0.8
1
0
0.5
1
1.5
2
2.5
3
3.5
4

**Figure 3**

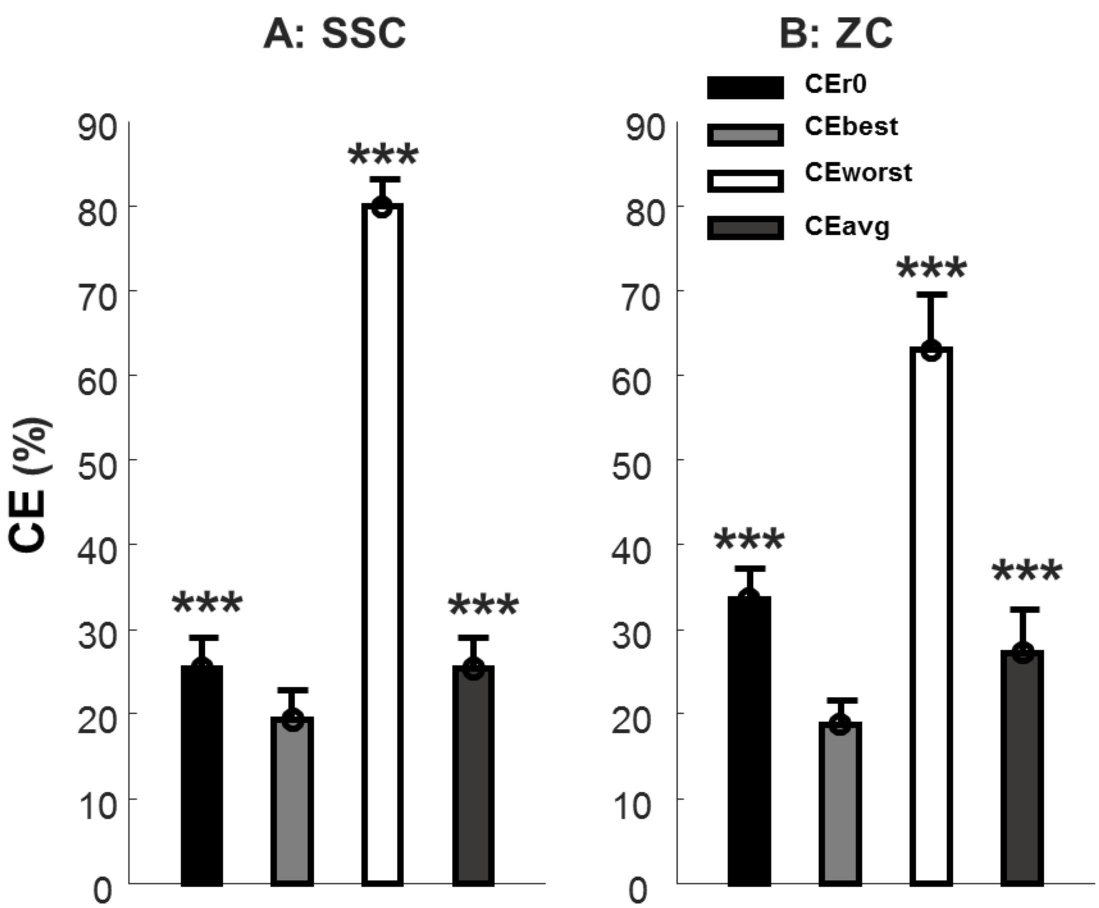

A: SSC
B: ZC
CEr0
CEbest
CEworst
CEavg
CE (%)
90
80
70
60
50
40
30
20
10
0
***

**Figure 4**

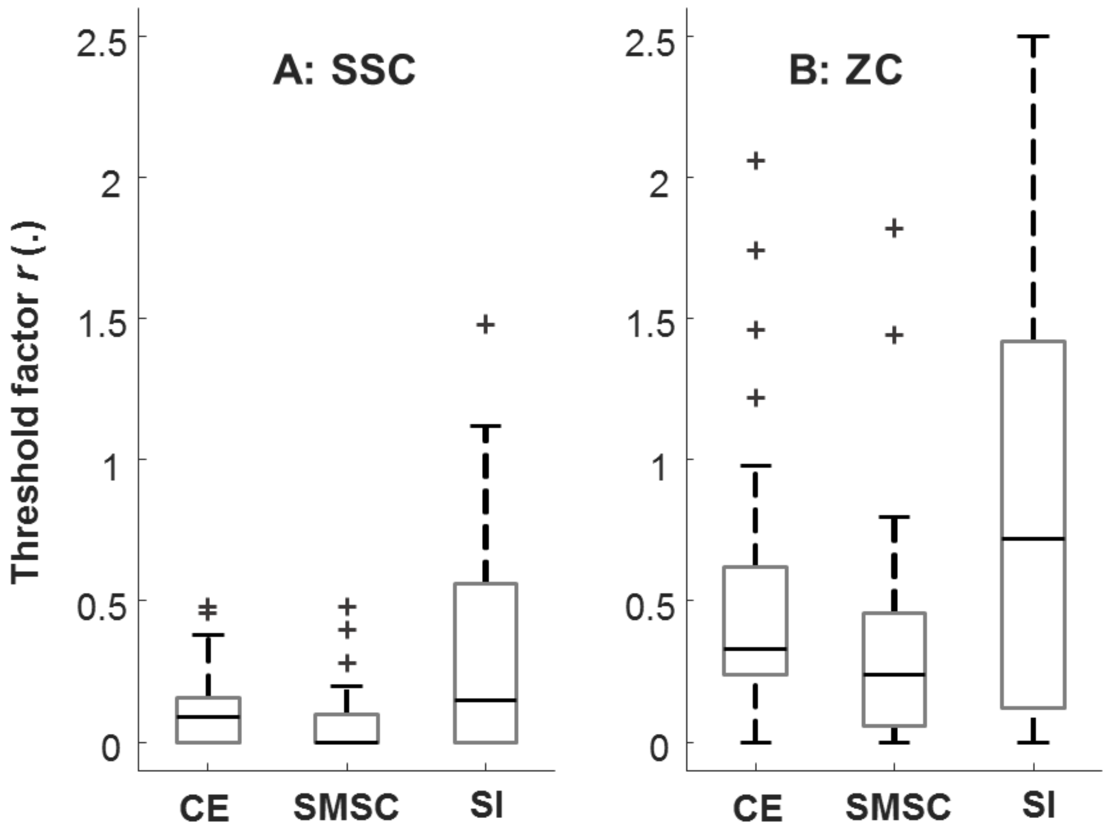


**Figure 5**

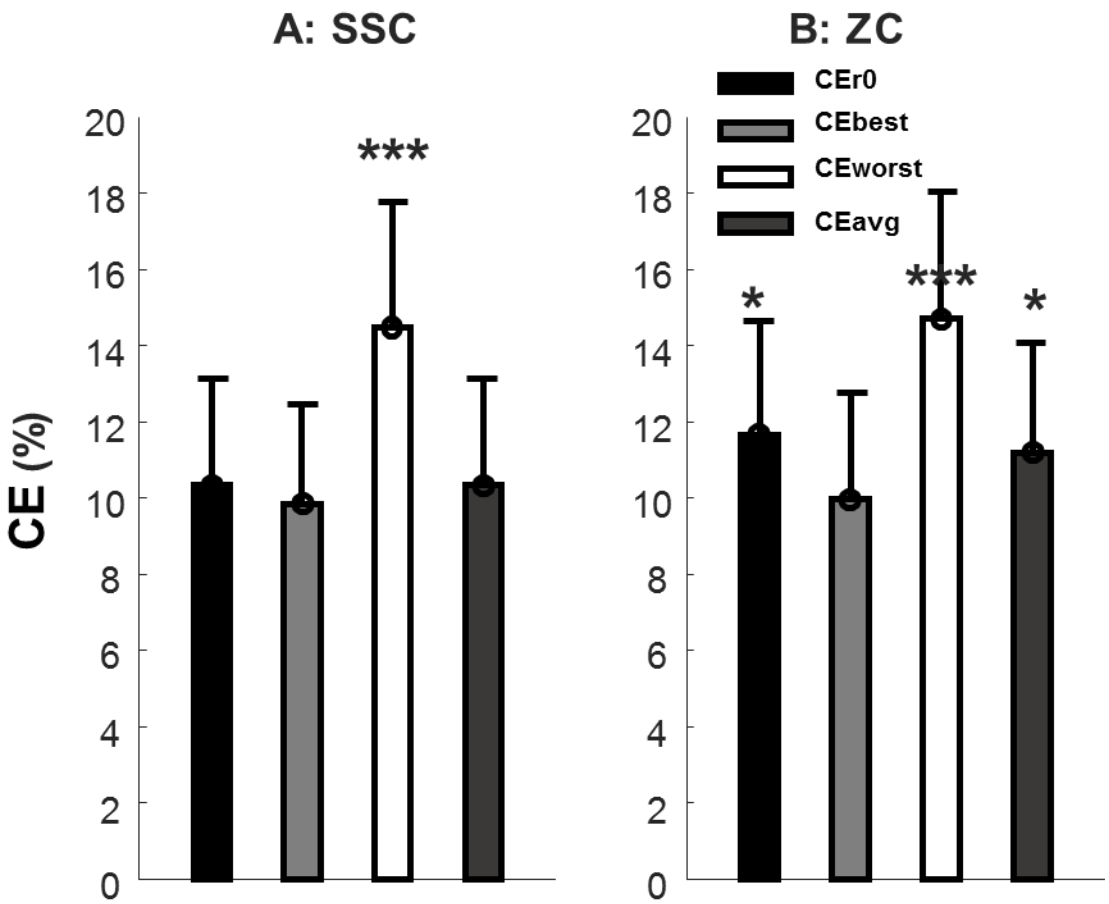


**Figure 7**

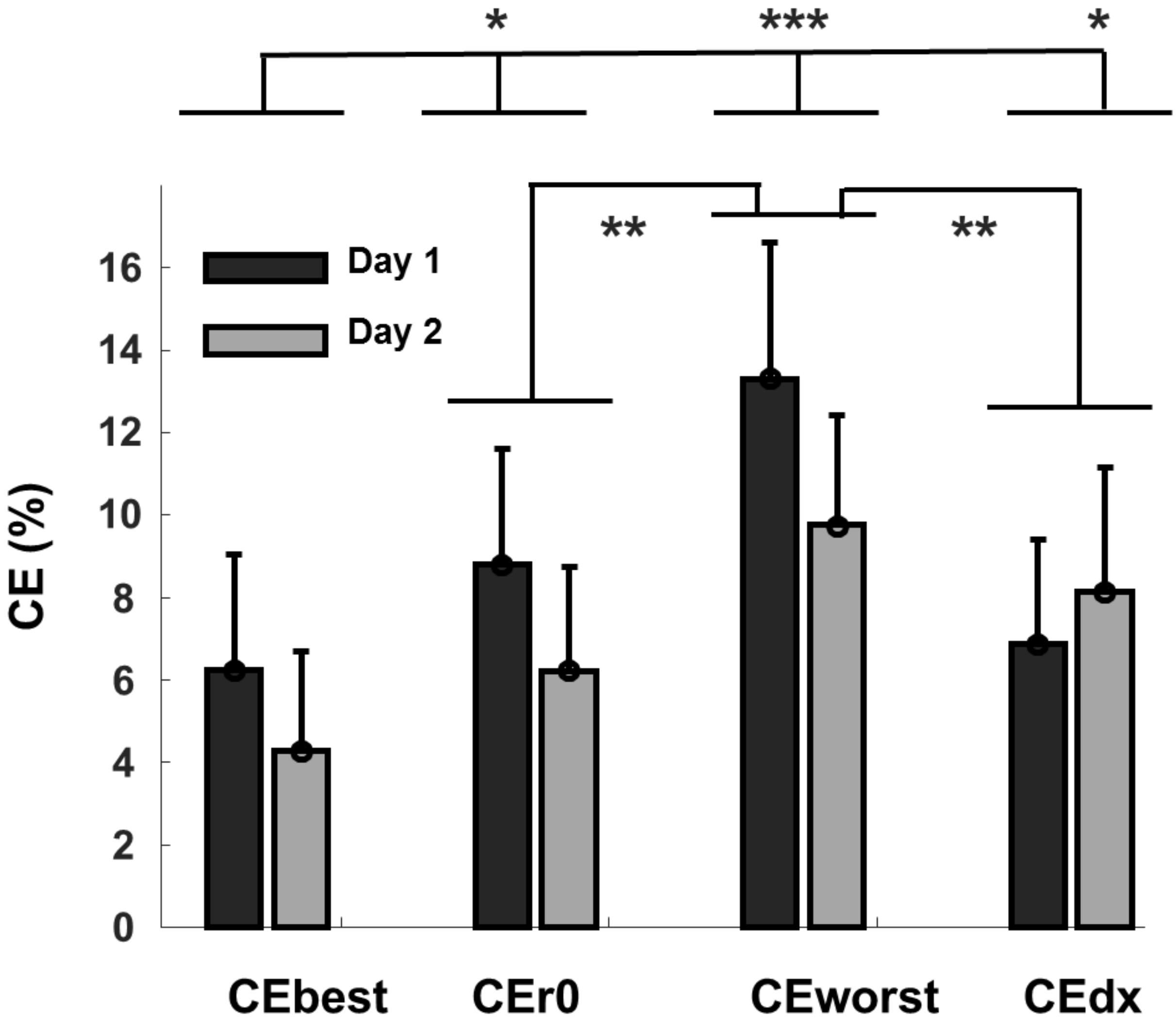
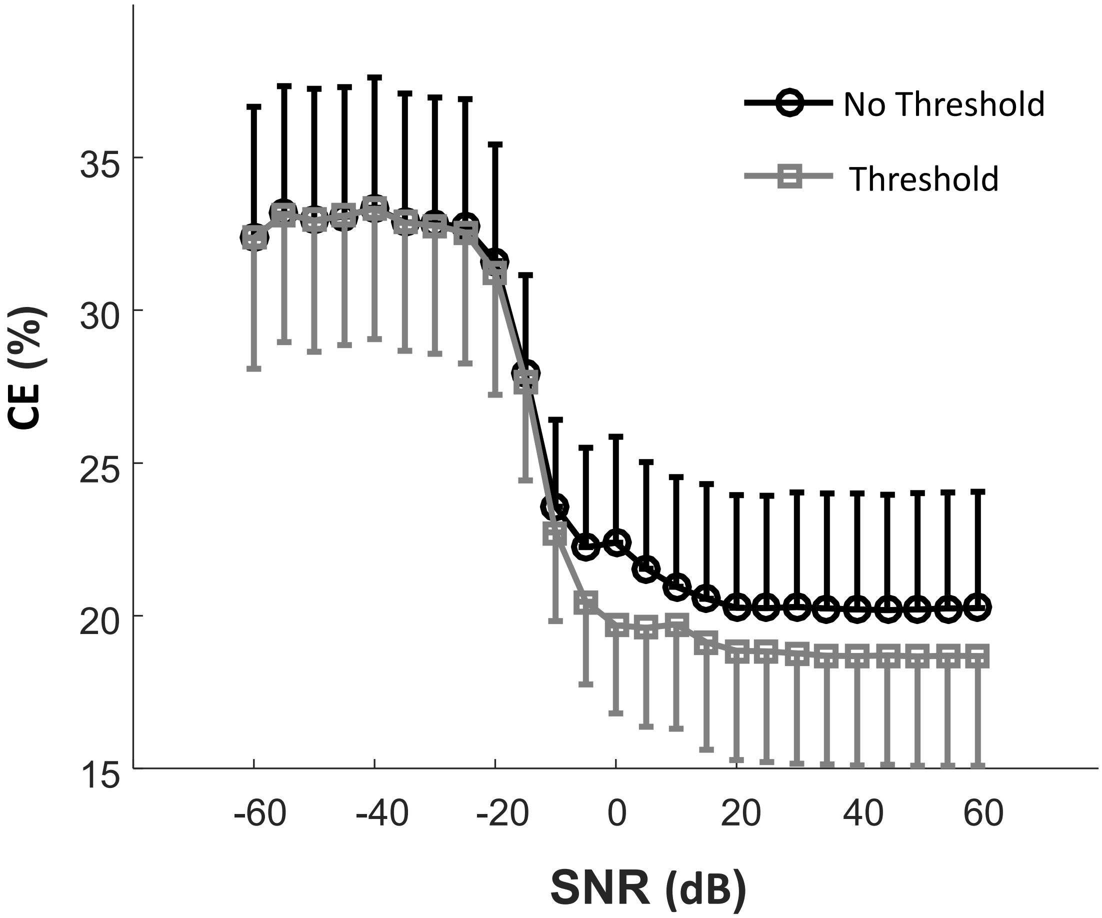


**Figure 6**


No Threshold
Threshold
35
30
25
20
15
CE (%)
-60
-40
-20
0
20
40
60
SNR (dB)